DOCUMENT RESUME

ED 045 959 AL 002 500

AUTHOR	Hays, Harriet Nutt
TITLE	From Mouth to Hand: Obstacles in Rendering Verbal Events Faithfully into Standard Orthography. The Classroom Interaction Project Series.
INSTITUTION	Missouri Univ., Columbia. Center for Research in Social Behavior.
REPORT NO	TR-CI-40
PUB DATE	24 Jul 70
NOTE	35p.; Paper presented at the Linguistic Society of America Annual Meeting, Columbus, Ohio, July 24, 1969
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Experiments, Graphemes, Interaction, Interaction Process Analysis, *Orthographic Symbols, *Phonetic Transcription, Statistical Studies, Tape Recordings, Video Tape Recordings

ABSTRACT

This paper reports the initial phase of a series of experiments conducted on a large number of videotapes made for the purpose of analyzing public-school classroom interaction. The experiments originally aimed to product the most reliable, efficient and economic way of producing transcriptions which are sufficiently representative of the verbal events used for empirical research. Results tabulated thus far indicate that, particularly among nonlinguists, but also among linguists, transcriptions of the same event into standard orthography are apt to differ to a significant extent; that some of these differences may not be entirely predictable; and that it takes at least two iterations of post-editing of the transcript to get a reasonable orthographic representation of the event. It also appears that the more complicated the structures involved, whether they be social, semantic or grammatical, the more verifications or post-editings are needed to produce an accurate transcription. The optimum work increment, processor personality, training or sequencing is not yet determinable, but, especially for difficult passages, it is likely that pairs of judges in the final editings, working together with transcript and tape, will be more efficient than single judges left alone with their idiosyncratic prejudices, anticipations, hearing and experience. (Author/AMM)

CENTER FOR RESEARCH IN SOCIAL BEHAVIOR

University of Missouri-Columbia

Technical Report #CI 40
The Classroom Interaction Project Series

FROM MOUTH TO HAND:

Obstacles in Rendering Verbal Events Faithfully into Standard Orthography

by

Harriett Nutt Hays

From Mouth to Hand: Obstacles in rendering verbal events faithfully into standard orthography

(Paper presented at the annual meeting of the Linguistics Society of America, July 24, 1970, at Columbus, Ohio, by Harriett Nutt Hays, Research Associate, Center for Research in Social Behavior, University of Missouri, Columbia, Missouri.)

To a phonetician or a lexicologist who has had direct experience in the field with variations among observers who transcribe or analyze oral data, either live or from electronic recordings, it is a matter of fact that all observers do not hear the same utterance in the same way. Kurath and McDavid, in Linguistic Atlas discussions, have reported variations among trained field workers. The Swedish dialectician Ringaard found that perceptions of other dialects by trained phonologists were influenced by their own manner of speech. Lieberman has noted that a great deal of the linguist's perception of prosodic features is based on intuition or knowledge of grammatical structures. I. H. Paul, a psychologist, noted some time ago that listeners' recall of a verbal event was subject to considerable variation among observers. Recently, Gumperz has bemoaned the fact that it is very tedious to obtain a transcript of an oral event for sociolinguistic studies.

What emphasis there has been on observer discrepancies has generally concentrated on subwordal, or phonological phenomena, or on more abstract functional or semantic entities. Very little has been done to examine the actual extent and cause of observer discrepancies relative to the translation of an oral situation into standard orthography. The assumption in many fields seems to be that a written transcription will not vary extensively from the real event if it concentrates on representing just the uttered words and sentences of that event: it is the prosodic information which is apt to be distorted. As we all have observed, transcriptions of wordal events are sometimes used as primary data for that event, for research, for legal action, for political record.

The clue to the general ignorance of the difficulties involved in capturing what is actually said may be reflected in the problems of illustrating and assessing them. It is, indeed, extremely tedious to represent oral information in a permanent form which is easily accessible

for perusal, and it is even more tedious to analyze varying representations of an event.

Signals made simultaneously in several dimensions of one medium have to be telescoped into perhaps only one or two dimensions of another. What is represented, for instance, in speech by the quantitative indicators (prosodic features) of amplitude, frequency, rate, and duration, which occur simultaneously with the qualitative indicators (consonants and vowels) are represented in writing as two-dimensional graphic signs (punctuations marks) which usually appear sequentially to the qualitative symbols (letters). The only simultaneous indicators for standard formal orthographic representations are capital letters, italics, underlines, boldface print, and the like. There is not a one to one ratio of the two symbolic systems, so their confusion is inherent in any transliteration. This is then compounded by the requirements of written discourse that all utterances be segmented by terminals which enclose strings of supposedly specific structures including what are referred to in standard school grammars as 'subjects,' 'predicates' and 'complete thoughts.' Spoken discourse, particularly with informal style, is characterized by what would be considered 'fragmentation' in written grammatical tradition. The problem of representing these 'incomplete thoughts' is difficult for translators who have been given no guidelines, particularly when they may have differing views of 'completeness' or of 'grammaticality.'

There are other problems which accompany the conversion procedure. Both the media and the situations for production of speech and production of its transliteration are different, presenting numerous possibilities for distortion. Speech, which can be considered the primary data, is a relatively unpredictable string of events incorporating a number of mutually interactive semiotic systems among which oral gesture and other expressive behavior patterns are included. Transliteration of the speech event is much more restricted in its potential boundaries than is the production of that event. Yet in both situations the participants have to add their individual interpretations to the events. Both are translation systems of a sort: speech is the translation into sound, apparently of thought product or behavioral convention; transliteration is a translation of a part of the speech event into a graphic medium. Just as speech suffers from limitations in its signals which do not

represent all the elaborations of the mind, so transliteration suffers from a lack of signals which simultaneously reinforce, extend or contradict nuances of the verbal message. The system which the transliterator uses seems rather to be designed to create graphic events such as essays, novels, poems or letters than to convert an alien system to its form.

The transliterator has the freedom of neither the author nor the speaker. Like the hearer and the reader, he must bring to the communication system his own experience in order to interpret the multi-referential signals which are employed. But unlike the hearer and the reader, he may not include them overtly in his transcription. His is a translation problem, in which a part of the other code is systematically left out. The results of such a conversion, without supplementation of electronic recordings of the oral event may be quite unusable as primary data, even though they may be more valid than the recall of the event by any one individual or individuals who do not commit the recall to a transcription.

This paper reports the initial phase of a series of experiments conducted on a large number of videotapes made for the purpose of analyzing public school classroom interaction. The original aim of the experiments was to predict the most reliable, efficient and economic way of producing transcriptions which are sufficiently representative of the verbal events to be used for empirical research.

The results of work tabulated thus far indicate that, particularly among non-linguists, but also among linguists, transcriptions of the same event into standard orthography are apt to differ to a significant extent, that some of these differences may not be entirely predictable, and that it takes at least two iterations of post-editing of the transcript to get a reasonable orthographic representation of the event. It appears also that the more complicated are the structures involved, whether they be social, semantic or grammatical, the more verifications or post-editings are needed to produce an accurate transcription. The optimum work increment, processor personality, training or sequencing is not yet determinable, but, especially for difficult passages, it is likely that pairs of judges in the final editings, working together with transcript and tape, will be more efficient than single judges left alone with their idiosyncratic prejudices, anticipations, hearing and experience.

The following discussion illustrates the kinds of omissions or distortions which the transliterator is apt to make when he transcribes an oral event using the standard literary graphic code. Note that the discrepancies tabulated refer only to the differences within the standard graphic transliteration system. They do not take account of the real situation, the actual amount of information of the communication system which is preserved or lost because of deficiencies in the system or elsewhere.

In order to assess the results, of course, it was necessary to use a preserved oral event. It was not possible at the time to make recordings specifically for this purpose and to have multiple observers on the spot judge the relative fidelity of the recorded material to the live situation. Such information was not extant for the recordings which were available, so no judgments about their actual fidelity to the live situation will be pertinent for this study.

It is pertinent, however, to know the conditions under which the recordings were made. Both segments of tape discussed in this study were recorded on the same 2400 foot roll of 2 inch 3M videotape, in the same urban elementary school, one each from two different sixth grade classes. Both of the teachers of these classes were young (20-30) females, who appear to speak the standard (prestige) teacher dialect of that Missouri city. The teacher of Segment I was black, of II was white. Class I was in English composition. Class II was in Social Studies, apparently a geography lesson in which a certain amount of reading aloud from the textbook took place. The pupils were male and female, black and white, children of approximately 12 years of age who spoke the local urban dialect but did not appear to have adopted so-called standard American.

Six microphones were used. The teacher wore a microphone suspended about the neck. The audio channel from this microphone was recorded on one track of a two track recording system. Four other microphones were hung from the ceiling and were recorded on the other track. All these were supplemented by another, directional, microphone which was aimed at the immediate emitter source. Two cameras were installed at opposite points in the classroom. The teacher was kept in constant focus, and the teacher picture inserted into an unused portion of the general classroom picture.

A detailed description of the recording equipment is found in Biddle and Adams, 1967. The recordings used for this study were made in 1968, for the Center for Research in Social Behavior at the University of Missouri, Columbia, Missouri. There was no special attempt made by teacher or pupils to enunciate or otherwise distort their behavior in order to improve the video and audio clarity of the recording. There were occasions of single as well as multiple responses by members of the class. The teacher was usually audible, but sometimes members of the class stationed far from the general microphones were difficult if not impossible to distinguish. Sometimes part of the class was not visible on the videoscreen. The lighting (unsupplemented on a rainy day), the distance, and the focus of the recording were such that the facial expressions of the class members were often not perceptible. Although a seating chart and rosters had been obtained at the time of recording, apparently there was no check on the actual position or presence of individual pupils, for neither these nor other tapes in the series reflect very well the arrangements indicated. (On some tapes there is no relationship at all between the rosters and the arrangement or content of the class). As we will see, this proves unfortunate, and, for those researchers for whom it is not a matter of course to diagram recorded events for location and activity of participants it would be well to take note. Literally hours of weeks have been spent by us attempting to straighten out boys from girls, black faces from white ones, high voices from low, etc., with very unsatisfactory results. It is our experience that, if the information is not gathered at time of collection, it might well be permanently lost. This loss of essential informant information then limits the use to which otherwise acceptable materials might be put.

The tape segments were arbitrarily chosen, with no pre-examination of the tape itself, for the contrasts originally were to serve as a quick illustration to associates that caution needed to be exercised in interpreting the oral material. The transcribers at the time were engaged in transcribing the tapes for the recorded series, and the next tape scheduled for transcription was selected for analysis.

Both transcribers and editors used the same playback equipment: a standard CONRAC 240" monitor set, an AMPEX VR01500 portable tape recorder with Play, Fast-Forward and Reverse controls handling two channels and accommodating a Tandberg 22 footpedal with two button controls for Forward and Reverse. A Sharp headset connected to a switch-box for signals from either channel singly, or both channels stereophonically, to the earphones, was used by some transcribers and editors. Others preferred to listen without earphones. No attempt was made to standardize or control this, but a casual survey indicates that it is probably more efficient to use the earphones, which seem to cut off transcription environment interference noise. Some persons, however, complained of headache from the headsets, and it was assumed that for them it was more efficient to work without both headache and headsets.

A survey of business concerns regarding optimal work increment for persons operating dictaphones seemed to indicate that a twenty minute period might be optimal for transcribing. This was shocking to the secreteries involved who had been in the habit of spending a much longer increment. A compromise of about 40 minutes was finally settled upon for transcription sessions. Time allowable for paper insert, forward and reversing the tapes and examination of the video image were assumed to constitute proper rest periods within the 40 minutes. (Examination of manuscripts seems to indicate that transcriber efficiency decreases rapidly at the end of 60 minutes.) About half of the editors claimed that they didn't start getting efficient until they had been working for about an hour, so there is certainly divided opinion on work increment. Some of the editors also claimed they could do better transcriptions than the transcribers by working at long intervals by hand. (This has been a common reaction. Most persons who have seen the lists of discrepancies have volunteered the information that they themselves would not make errors.) Since none of them turned out to be infallible, it is still unclear whether relative time has any bearing on the quality of output. There is no question about the fact that it is much more expensive to employ a slow working than a fast working editor.

Editor is really a misnomer. The editors were not to function as standard literary editors do. Rather they were to reproduce an utterance without improving upon it. They were to compare all previous graphic versions against the tape of the situation and to insert, delete or modify

those translations which seemed inappropriate. Similarly, the transcriber was to reproduce rather than beautify the original utterance.

The regular transcribers for the interaction study then, four secretaries employed by the research center, one a college graduate, the others with high school degrees and secretarial training, transcribed Class I using the conventions they had already established: All senders of utterances were designated in parentheses, left-justified on a new line. All public utterances were transcribed into standard orthography. Indistinguishable utterances were represented by a line, whose length might or might not have an impressionistic relationship to length of utterance.

Class II was transcribed by the same set of secretaries but after a two hour session to establish additional conventions. Annotations or disambiguations to clarify the context of ambiguous utterances were to be inserted in slashes, on the assumption that, if they were properly marked they could easily be left out. Punctuation was to be reduced to a minimum, where possible. For terminals, only period, indicating a statement neutral in tone or feeling, question mark to indicate a definite question contour, and three dots to mark a suspended or unfinished oral sentence were to be used. Quotes were to enclose matter being read out loud by an emitter. (The later convention ...? used by some editors, had not yet been established. The practice of indicating pauses was also not standardized until the post-editing was well under way.)

The determination of the length of the tape segments contrasted was calculated from the first phrase for which there was consensus among versions to the last event of the shortest transcript, a total, for Class I, of seven minutes of real time. The initial segment, on which this common point was established is represented on page one of the handout. The shortest transcript was that of c, which had one indistinguishable utterance of the first emitter before the common beginning phrase and none at the end of the common transcription. Next was d with two distinguished utterances by the first emitter, before the common starting point, and 4 lines (2 emitters) beyond the end point. B began at the common point, but continued for 9 lines (6 emitters) past the common end. A had one distinguishable utterance before the first, and 11 lines (6 emitters) after the end point of the common transcription.

The corresponding transcription segments were compared for gross characteristics which are often used to describe manuscripts: number of lines, sentences, words, emitter types and word totals. Page and line are obviously inadequate descriptive categories for manuscripts unless they are all made on the same size paper with the same size type. The typed abcd transcripts were, but some of the other transcripts in this study differed in line spacing, size of paper, and size of graphic symbols. (The handwritten copies had either bigger or smaller symbols than those which were typed.)

Number of sentences for the manuscripts was similar, but on close examination the content of the terminals were found to contrast sharply. In the tally for potential sentences, 91 potentials, or 4 more than the greatest number of sentences indicated by any one transcriber were postulated, but might vary for each analyst contrasting the scripts.

The fifth original transcript was made by secretary b, an arbitrary choice, for both Classes. All transcripts but II c and d were then edited by a team of individuals listed on page 8 of the handout.

The first four unedited or raw transcripts for Class I were extensively contrasted with each other for differences in major block categories of Emitter, Annotation, Punctuation and Utterance. These were subdivided further according to a hierarchial code devised to prepare the data for ultimate input into a computer, where the long lists of idiosyncratic and other deviations might be tallied with greater ultimate ease, or at least accuracy.

From the raw transcripts for Class I a handwritten transcription (E) was calculated based upon majority agreement of processors and contextual fit. This calculation was done by a naive editor, that is to say a non-linguist, relatively unsophisticated female white sophomore. It was edited against the tape by the same person, then re-edited by myself and a sophisticated, acute, male white sophomore. The second transcription by b, (F) was edited four times against the tape. An exhaustive chart was drawn to align the manuscripts, and make some hand tabulations, by a linguistically naive female white senior, with apparently good judgment.

For some of the contrasts, more than one chart was drawn and compared. G, m, and n all worked on contrasting some of the versions.